Climatic Change
DOI 10.1007/s10584-013-0990-2

A framework for testing the ability of models to project climate change and its impacts

J. C. Refsgaard · H. Madsen · V. Andréassian · K. Arnbjerg-Nielsen · T. A. Davidson · M. Drews · D. P. Hamilton · E. Jeppesen · E. Kjellström · J. E. Olesen · T. O. Sonnenborg · D. Trolle · P. Willems · J. H. Christensen

Received: 31 March 2013 / Accepted: 14 October 2013

Abstract Models used for climate change impact projections are typically not tested for simulation beyond current climate conditions. Since we have no data truly reflecting future conditions, a key challenge in this respect is to rigorously test models using proxies of future conditions. This paper presents a validation framework and guiding principles applicable across earth science disciplines for testing the capability of models to project future climate change and its impacts. Model test schemes comprising split-sample tests, differential split-sample tests and proxy site tests are discussed in relation to their application for projections by use of single models, ensemble modelling and space-time-substitution and in relation to use of

J. C. Refsgaard (✉) · T. O. Sonnenborg
Geological Survey of Denmark and Greenland (GEUS), Øster Voldgade 10, 1350 Copenhagen K, Denmark
e-mail: jcr@geus.dk

T. O. Sonnenborg
e-mail: tso@geus.dk

H. Madsen
DHI, Agern Alle 5, 2970 Hørsholm, Denmark
e-mail: hem@dhigroup.com

V. Andréassian
IRSTEA, Antony, France
e-mail: vazken.andreassian@irstea.fr

K. Arnbjerg-Nielsen
Technical University of Denmark, 2800 Lyngby, Denmark
e-mail: karn@env.dtu.dk

T. A. Davidson
Aarhus University, Vejlsøvej 25, 8600 Silkeborg, Denmark
e-mail: thd@dmu.dk

M. Drews
Technical University of Denmark, Risø, 4000 Roskilde, Denmark
e-mail: mard@dtu.dk

Published online: 01 November 2013

different data from historical time series, paleo data and controlled experiments. We recommend that differential-split sample tests should be performed with best available proxy data in order to build further confidence in model projections.

1 Introduction

It is good practice in many earth science disciplines to conduct rigorous testing of models before they are applied for predictions (Jørgensen 1995; Refsgaard and Knudsen 1996; Wolf et al. 1996). Such tests may be denoted validation tests, history matching or similar depending on the terminology used. Climate change poses a fundamental new challenge as the models are used for projections of an unknown future with a climate significantly different from the current conditions. Almost all model studies reported in the scientific literature apply models to make projections of climate change and its impacts without any prior assessment of the credibility of the models for simulation beyond current climate conditions (Wagener et al. 2010). This is a significant weakness that implies an unknown level of uncertainty of the projections. The key challenge is that we cannot, in a strict sense, perform validation tests on the ability of our models to project the climate change effects since we have no data truly reflecting the future conditions.

In pioneering work, Klemes (1986) proposed a systematic testing scheme designed for assessing the feasibility of using models for simulating conditions under a changed climate. While Klemes' validation tests or variations thereof have been applied in hydrology (Seibert 2003; Coron et al. 2012), they have not yet found widespread use in other earth science disciplines.

The objectives of this paper are to develop a validation framework and guiding principles applicable across earth science disciplines for testing the capability of models to project future climate change and its impacts.

D. P. Hamilton
Environmental Research Institute, University of Waikato, Hamilton, New Zealand
e-mail: davidh@waikato.ac.nz

E. Jeppesen · D. Trolle
Department of Bioscience, Aarhus University, Vejlsøvej 25, 8600 Silkeborg, Denmark

E. Jeppesen
e-mail: ej@dmu.dk

D. Trolle
e-mail: dtr@dmu.dk

E. Kjellström
Swedish Meteorological and Hydrological Institute, Norrköping, Sweden
e-mail: erik.kjellstrom@smhi.se

J. E. Olesen
Aarhus University, 8830 Tjele, Denmark
e-mail: JorgenE.Olesen@agrsci.dk

P. Willems
KU Leuven, Leuven, Belgium
e-mail: Patrick.Willems@bwk.kuleuven.be

J. H. Christensen
Danish Meteorological Institute, Lyngbyvej 100, 2100 Copenhagen Ø, Denmark
e-mail: jhc@dmi.dk

2 Framework

2.1 Terminology

Our terminology is inspired by Schlesinger et al. (1979) and Refsgaard and Henriksen (2004). The modelling environment is divided into four basic elements as shown in Fig. 1. The inner arrows describe the processes that relate the elements to each other, and the outer circle refers to the procedures that evaluate the credibility of these processes.

In general terms, a model is understood as a simplified representation of the natural system that it attempts to describe. However, a distinction is made between three different meanings of the general term model:

- *conceptual model*—the model structure and equations;
- *model code*—the model software that is selected or constructed to allow numerical implementation of the conceptual model;
- *site and context specific model*—a set-up of the model code for a given domain, time period and purpose, including the input data, model forcing and parameter values.

All three model types are subject to testing, which is conditional on certain specified limits of model application and corresponding levels of accuracy:

- *confirmation* implies determination of the adequacy of the *conceptual model*, i.e., the scientific confirmation of the theories/hypotheses included in the conceptual model;
- *verification* implies substantiating that a *model code* provides a sufficiently accurate solution of the equations and structure represented in the conceptual model;
- *validation* implies substantiating that a *site and context specific model* can project future conditions with a satisfactory level of accuracy.

The *model confirmation and validation* processes are particularly challenging in a climate change impact context because proxy data are required due to the lack of data on future conditions (Fig. 1).

In many earth science disciplines model *calibration* is performed as part of the model set-up to optimise the parameter values in such a manner that the model output best fits field observation data.

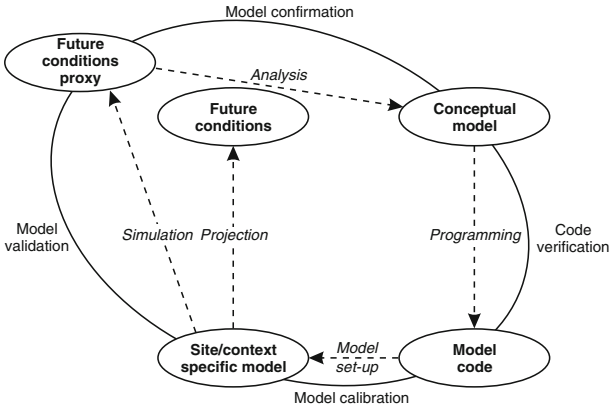

Fig. 1 Elements of a modelling terminology. The *dashed lines* represent processes that relate the four basic elements to each other, while the *full line curves* refer to the procedures that evaluate the credibility of these processes. Modified from Refsgaard and Henriksen (2004)

From a scientific philosophical point of view verification and validation of numerical models of natural systems are impossible, because natural systems are never closed, and hence the mapping of model results is never unique (Popper 1959; Oreskes et al. 1994). Therefore, it is not possible to carry out model verification or model validation if these terms are used universally without restriction to domains of applicability and levels of accuracy. Popper (1959) distinguished between two kinds of universal statements: the 'strictly universal' and the 'numerical universal'. The strictly universal statements are those usually implied when discussing theories or natural laws. In contrast, numerical universal statements refer only to a finite class of specific elements within a finite spatio-temporal region.

According to Popper's views the restrictions in use of the terms confirmation, verification and validation, imposed by the respective domains of applicability, imply that the conceptual model, model code and site and context specific models can only be classified as numerical universal statements and not as strictly universal statements. This distinction is fundamental for our terminology. Hence, a model can never be universally validated; rather the validation is conditional on specific locations, acceptable accuracy/uncertainty levels and types of application.

2.2 Projection methodologies

Due to the stochastic nature of weather systems climate models cannot provide predictions of specific future weather events beyond 1 to 2 weeks. Instead their aim is to provide information on the statistical properties of the future climate under a given scenario. This is denoted model *projections* (Taylor et al. 2012). As downstream elements of the modelling chain, such as hydrological and ecological models, operate on outputs from climate models, they are also confined to making projections rather than predictions (see Fig. 1).

Methodologies used for making projections on future climate and its impacts may be classified into three types (Fig. 2):

- *Single model:* Projections based on a single model.
- *Model ensemble*: Projections based on an ensemble of different models, including different model codes, forcings and parameterisations.

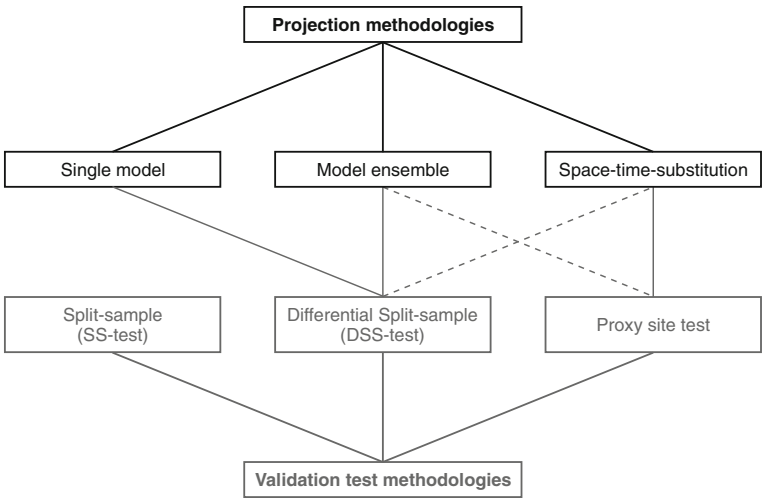

Fig. 2 Classification of methodologies for testing a model's capability to project climate change effects

- *Space-time-substitution*: Identification of a place or a number of places having a past or current climate similar to the projection of the future climate at the site of interest and use of data from these places as a proxy for the impacts of the projected climate change.

2.3 Validation methodologies

Testing schemes for validation need to reflect site and purpose specific conditions including the type of model and the data availability, and hence the implementation of the schemes will differ from case to case. The guiding principles required to achieve this goal are:

- Before it is used operationally, a model must demonstrate how well it can perform the kind of task for which it is intended. As a model is aimed at providing insight into the unknown future, it cannot be tested on data for such a situation but rather on data pertaining to a situation similar to the expected future situation (Klemes 1986).
- The validation test must be carried out against independent data that have not been used for model calibration (Klemes 1986).
- The validation test must provide information on the expected accuracy of the model projections (Refsgaard and Henriksen 2004; Christensen et al. 2010).

Validation methodologies may in accordance with Klemes (1986) be classified into three types (Fig. 2):

- *Split-sample test (SS-test)*: The data are, temporally and/or spatially, split into two parts. One part is used for calibration, while the other part is reserved for validation. A more sophisticated version of the SS-test is jack-knifing where the data split and testing are repeated systematically, so that all data are used for both calibration and testing. The underlying assumption behind the SS-test is that climate conditions, as well as physical conditions, can be assumed stationary.
- *Differential split-sample test (DSS-test):* DSS-tests are applicable if climate conditions are non-stationary. The test implies that a model ideally is tested against observation data similar to the future climate conditions. Due to lack of such data, DSS-tests are often made using periods with apparent different climate conditions (e.g. dry/wet or cold/warm) where calibration is performed on one period and validation on another period.
- *Proxy site test:* A proxy site test is a test of the capability of the model to project conditions without prior calibration, i.e. completely without calibration or by calibration at some locations with site-specific data and projection at the location of interest.

3 State-of-the-art projection and validation methodologies

Table 1 provides an overview of the use of projection methods, validation tests and data sources within different disciplines in the climate impact modelling chain.

3.1 Projection methodologies

Using a *single model* is the simplest method for making projections. This has been state-of-the-art in most impact studies in hydrology, agroecology and freshwater ecology until a few years ago (Bates et al. 2008; Challinor et al. 2009; Trolle et al. 2011). Typically, the use of single models has been associated with model calibration ensuring that the model performance for the present period can be tested.

Table 1 Level of use of projection methods, validation test methods and data sources in the modelling chain

	Climate modelling	Statistical downscaling	Hydrology	Agroecology	Freshwater ecology
Projection method					
Single model	XX	XXX	XXX	XXX	XXX
Ensemble models	XXX	XXX	XX	XX	X
Space-time-substitution			X	X	XX
Validation test					
SS-test	XX	XX	XXX	XX	XX
DSS-test		X	XX	X	
Proxy site test	XX		X	X	X
Data source					
Historical time series	XXX	XXX	XXX	XXX	XX
Paleo data	X				X
Controlled experiments	X			X	X

XXX widespread use, *XX* some use, *X* few examples

Acknowledging the large differences between projections from different climate models, *ensemble modelling* has been standard among climate modellers for the last two decades (van der Linden and Mitchell 2009; Taylor et al. 2012). From different ensemble model experiments, it is generally concluded that the ensemble average provides a better representation of the current climate than any single climate model (Gleckler et al. 2008). The ensemble is therefore expected to provide a more robust projection of the future climate. In addition, it provides an estimate of the associated projection uncertainty. In impact studies, both for single elements and the entire modelling chain, ensemble modelling is gradually being accepted as state-of-the-art (Wilby and Harris 2006; Rötter et al. 2011). Models are often, but not always, calibrated. Experience has shown that even models that through calibration have been forced to achieve similar performance in a calibration period may exhibit quite different projections for future climates (Wolf et al. 1996; Velázquez et al. 2012).

Use of climate and impact models requires that these have the capability to project climate change and its impacts, i.e. that the underlying conceptual models include process descriptions enabling them to project how present conditions will be transformed in future climates. The status of the different disciplines (Table 1) reflects differences in model accuracy, data availability and associated scientific traditions. In some disciplines, for example freshwater ecology, the present knowledge is still so incomplete that the conceptual models often cannot be confirmed, making the basis for model projections very questionable (Trolle et al. 2011). In such situations *space-time-substitution* offers the possibility to use data from other areas with different climates to infer about climate change impacts in the area of interest. Space-time-substitution is used in freshwater ecology (Meerhoff et al. 2012) and has also been attempted in hydrology (Singh et al. 2011) and agroecology (Elsgaard et al. 2012).

3.2 Validation methodologies

SS-tests are in many cases the only type of test used, probably because it is the easiest and the only test type for which data are readily available. As the underlying assumption behind the SS-test is that climate conditions are stationary, it is not an adequate test in climate change impact studies. This conclusion is confirmed by the results from DSS-tests described below.

DSS-tests have been used for both single models and ensemble models. In a hydrology study Refsgaard and Knudsen (1996) used three models for two catchments in Zimbabwe, calibrating on wet periods and validating on dry periods. Seibert (2003) used a model to simulate peak flows for four Swedish catchments, calibrating on periods with small peaks and validating on periods with large peaks. Coron et al. (2012) applied a more sophisticated DSS scheme with three models and data from 216 catchments in Australia. Their key finding (Fig. 3) confirms the conclusions of other studies: that model performance generally deteriorates when making predictions for other periods than the calibration periods, and that the loss of performance increases the more the climate conditions in the prediction period differ from the conditions in the calibration period.

A fundamental assumption in statistical downscaling of climate model projections is that the climate model biases are stationary. This assumption is questionable, and recent research shows that model biases may not be stationary in a warming climate (Boberg and Christensen 2012). Thus, it is important to evaluate the robustness of the statistical downscaling methods with respect to a non-stationary bias. To date, however, validation tests have not been widely applied in this respect. Recently, Teutschbein and Seibert (2012) applied a DSS-test to evaluate different bias correction methods for downscaling an ensemble of regional climate model projections. They found that the simpler correction methods, such as the widely applied delta-change approach, are less robust to a non-stationary bias compared to more advanced correction methods.

Years with extreme or contrasting weather conditions have recently been used to assess the ability of crop models in simulating yield responses to variability in the weather (Eitzinger et al. 2012). The results showed that even a short period of extreme hot weather may cause considerable differences among models.

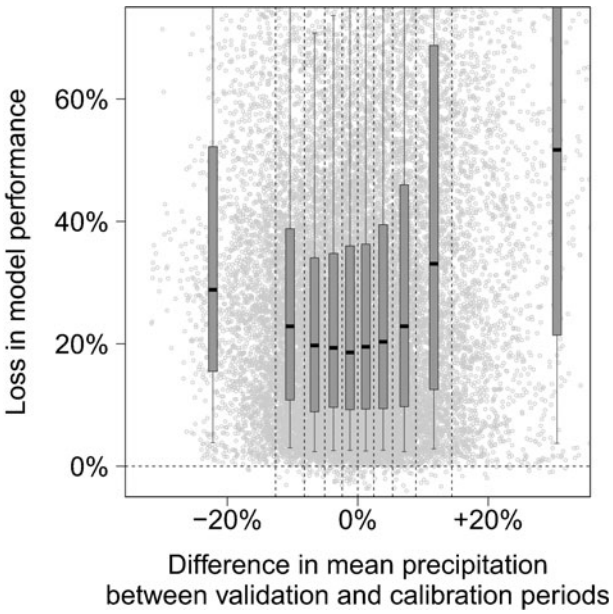

Fig. 3 Relative loss of performance (y-axis) against relative evolution of climate conditions (x-axis) for ensemble split sample tests using three hydrological model codes on 216 catchments in Australia. Each *dot* represents a DSS test for one catchment with one combination of calibration/validation periods and one model, while the *box plots* show the 0.05, 0.25, 0.5, 0.75 and 0.95 percentiles. Modified from Coron et al. (2012)

We are not aware of examples where DSS-tests have been used together with space-time-substitution.

Proxy site tests are carried out when models are not calibrated but subject to testing against field data in the present climate. This is common practise for climate models, where data for DSS tests are not available. In other disciplines it is less used (Table 1). However, it has been applied for prediction in ungauged basins in hydrology (Refsgaard and Knudsen 1996) and for testing a range of crop simulation models with explicit restrictions on model calibration (Palosuo et al. 2011; Rötter et al. 2011). As space-time-substitutions use data from other locations or periods, any test of this projection methodology involves elements of proxy site tests.

3.3 Data sources for validation tests

The most commonly used data source within all disciplines is data from the recent past climate. This provides possibilities to test the capability of models to reproduce climate variability and its impacts as well as spatial differences. In order to perform more powerful DSS-tests data sets containing non-stationarity are desirable. In this respect three data sources are interesting: (i) historical time series comprising non-stationarity, (ii) paleo data and (iii) data from controlled experiments.

Many historical time series exhibit non-stationarity, mostly due to human activity such as land use change or river regulation. An example where climate is the dominant source of non-stationarity is the time series from Skjern River in western Denmark where the precipitation has increased by 26 % and the temperature by 1.3 °C since 1875 (Karlsson et al. 2013). A DSS-test showed that a hydrological model had difficulty in predicting the changes in runoff during this transient period.

As suitable historical time series containing non-stationary signals are limited, palaeolimnological methods tracking long-term climate change beyond the scope of the instrumental record may be used (Olsen et al. 2012). In keeping with climate change projections, models for tracking past precipitation have proved more problematic than those for temperature. There is, however, great potential, perhaps through closer integration of modelling and palaeolimnological approaches, in validating model predictions by comparing current change with past ecological responses during periods of known cooling or warming (e.g. the little ice age, medieval warm period).

Controlled experiments in the field or laboratory have also been used to elucidate and quantify the ecosystem effects of climate change (Jentsch et al. 2007). Since such experiments, as for numerical models, represent simplifications of the real world, the signals they produce can help evaluate the validity of the model projections. As researchers may both design the experimental environment and control the external forcing, the responses to climate forcing can be studied at system/ecosystem scale (e.g. Liboriussen et al. 2011) as well as for specific processes occurring within the system (Jensen and Andersen 1992), allowing intensive scrutiny of the numerical model projections.

4 Discussion

4.1 Towards an improved practice of validation tests

It is good scientific practice to test one's hypothesis before applying it in practice. This is, however, difficult to honour completely in connection with climate change impact studies,

because data on the conditions prevailing under a future climate generally do not exist. We have proposed guiding principles and a classification of test schemes for evaluating a model's capability to perform climate change impact projections. We argue that the most commonly used SS-test is inadequate and that the DSS-test is more appropriate. DSS-tests are, however, not commonly used in any of the earth science disciplines. This may be due to lack of tradition, as such a test type can be used in many more situations than today. It is reasonably straightforward to use DSS-tests on climate variability data in historical time series (Seibert 2003; Coron et al. 2012; Fig. 3). Although such tests are not full DSS-tests, insofar as the future climate will often represent conditions that are beyond the conditions found in observation data, we argue that DSS-tests are the best possible evaluation method.

Another example where DSS-tests may be introduced is in connection with space-time-substitution projections, for which validation tests are usually not performed. For many designs of space-time-substitution schemes it would be possible to add a DSS-test to evaluate how well such projections are able to reproduce temporal variability at different locations. To our knowledge the only example where this has been carried out is the study by Singh et al. (2011) who first regionalised climate dependent streamflow characteristics from 394 catchments in the USA and then assumed that this spatial relationship between climate and streamflow characteristics is similar to the one that would occur under a future climate change. They subsequently tested this assumption for five catchments by use of five 10-year historical datasets, where they calibrated on one period and made DSS-tests on four other periods with different precipitation characteristics.

Altogether, we strongly recommend that DSS-tests should be performed with best available proxy data in order to learn and build further confidence in model-based climate change projections.

4.2 Ensemble representation

Validation tests are generally expected to be more powerful when applied to a model ensemble than to a single model. However, a particular problem related to ensemble modelling is that the models in the ensemble often are not mutually exclusive and collectively exhaustive (Bommer and Scherbaum 2008), i.e. they are dependent and do not cover the full space of plausible models. This is true for both multi-model ensembles due to the relatively low number of available models (van der Linden and Mitchell 2009) and perturbed physics ensembles (Murphy et al. 2007) as these do not sample structural model uncertainty. Sunyer et al. (2013) showed that the information content in an ensemble of 15 regional climate models from the ENSEMBLES project (van der Linden and Mitchell 2009) may correspond to as little as five equivalent independent models, depending on the climate statistic considered.

Due to model dependencies and common model biases the ensemble mean projection may be biased and the uncertainty represented by the ensemble will be underestimated. Thus, one may want to weight models according to performance and even exclude outlier models from the ensemble. Assignment of weights to the ensemble members may be a plausible way of incorporating model accuracy and credibility in the ensemble projection, but it may be difficult to apply in practice. Since model performance cannot be directly assessed for future conditions, additional assumptions are needed, such as assuming time invariant model bias or using a pseudo reality for assessing future model performance (such as the ensemble mean projection used by Tebaldi et al. (2005)). An additional practical problem is that models perform differently with respect to different performance measures (Christensen et al. 2010).

4.3 Validation in the modelling chain

We have so far focussed on validation of the individual elements of the modelling chain from global and regional climate models and statistical downscaling to hydrological, agro-ecological and freshwater ecological impact models. However, the accuracy of model projections in the upstream chain will influence the credibility of the projections of downstream impact models (Refsgaard et al. 2013). Although the impact models can be validated using data within their own modelling regime, biases and uncertainties introduced in the upstream part of the modelling chain may result in forcing of downstream models that is way out of the range for which the models have been validated. This emphasises the need for a careful validation of all parts of the modelling chain that includes key statistical properties of variables (e.g. average conditions, variability or extremes) sensitive to downstream model performance (Swayne et al. 2005). Ideally, all elements in the modelling chain should be validated in an integrated manner. As individual studies usually do not carry out modelling of all elements, this is difficult in practice, and we are not aware of studies that have performed integrated validation tests incorporating all elements in the modelling chain.

Although our guiding principles should be applicable to all parts of the modelling chain, it is important to acknowledge that uncertainties in observations and lack of knowledge of processes differ between the disciplines with a general tendency of larger uncertainties in biological variables compared to chemical variables, which in turn are more uncertain than physical variables. Thus, we may never be able to achieve large credibility in projections of some state variables despite following a rigorous validation protocol in all parts of the modelling chain. In this respect we also have to keep in mind that the models in the model chain are generally not dynamically coupled. Important feedback mechanisms may therefore be missing, potentially affecting the validation. Here, space-time-substitution, controlled experiments and use of paleo data can be useful supplementary methods to dynamic modelling as many of the uncertainties and potential non-linear relationships are implicitly included in the data.

4.4 Performance criteria

The performance criteria used in validation tests should reflect the conditions in climate change projections and be purpose-specific.

Climate models are not able to reproduce single events but rather statistical properties describing the climate. This should also be reflected in tests of, for instance, hydrological and agroecological models. Therefore, commonly used performance criteria in hydrology and agroecology, such as the root mean square error and the Nash-Sutcliffe coefficient focussing on temporal fits to observed data, are not relevant for this type of test. Criteria based on flow duration curves or quantiles in probability distribution functions are much more appropriate. Similarly, in the case of ensemble model projections, performance criteria need to reflect that ensemble modelling is probabilistic, i.e. performance criteria should not be applied to individual models but rather to the uncertainty intervals represented by the ensemble. For example, a 90 % confidence interval should contain 90 % of the observations.

Purpose-specific performance criteria imply that different criteria should be applied for different study purposes (Madsen 2000). For climate change impacts on floods, for instance, it may be relevant to use performance criteria that measure how well models predict extreme floods, such as floods with recurrence intervals of 5, 10 or 50 years (van Steenbergen and Willems 2012). For water availability and ecosystem studies, it may instead be relevant to use performance criteria measuring the ability to reproduce mean annual flows or minimum flows.

For crop yield modelling a relevant criterion could be the frequency of yields below a critical level that exceeds the coping range of the particular farming system (Palosuo et al. 2011). Similarly, for lake modelling a relevant criterion could be the probability of exceeding a critical threshold for abundance of potentially toxic cyanobacteria.

One way of dealing with this issue is already in the calibration process to carefully select objective functions for the parameter optimisation that are purpose specific and reflect the most important statistical properties of the model projection. Van Steenbergen and Willems (2012) proposed a calibration methodology that focuses on the factors controlling changes in peak flow as a response to changes in climate conditions and showed that this provided more reliable predictions of extremes.

Acknowledgments The present study was funded by a grant from the Danish Council for Strategic Research for the project Centre for Regional Change in the Earth System (CRES—www.cres-centre.dk) under contract no: DSF-EnMi 09-066868.

References

Bates BC, Kundzewicz ZW, Wu S, Paulikof JF (eds) (2008) Climate change and water. Technical paper of the intergovernmental panel on climate change. IPCC Secretariat, Geneva

Boberg F, Christensen JH (2012) Overestimation of Mediterranean summer temperature projections due to model deficiencies. Nat Clim Chang 2:433–436

Bommer JJ, Scherbaum F (2008) The use and misuse of logic trees in probabilistic seismic hazards analysis. Earthquake Spectra 24(4):997–1099, Earthquake Engineering Research Institute

Challinor AJ, Ewert F, Arnold S, Simelton E, Fraser E (2009) Crops and climate change: progress, trends and challenges in simulating impacts and informing adaptation. J Exp Bot 60:2775–2789

Christensen JH, Kjellström E, Giorgi F, Lenderink G, Rummukainen M (2010) Weight assignment in regional climate models. Clim Res 44(2–3):179–194

Coron L, Andréassian V, Perrin C, Lerat J, Vaxe J, Bourqui M, Hendrickx (2012) Crash testing hydrological models in contrasted climate conditions: An experiment on 216 Australian catchments. Water Resour Res 48, W05552

Eitzinger J, Thaler S, Schmid E, Strauss F, Ferrise R, Moriondo M, Bindi M, Palosuo T, Rötter R, Kersebaum C, Olesen JE, Patil RH, Saylan L, Caldag B, Caylak O (2012) Sensitivities of crop models to extreme weather conditions during flowering period demonstrated for maize and winter wheat in Austria. J Agric Sci (in press)

Elsgaard L, Børgesen CD, Olesen JE, Siebert S, Ewert F, Peltonen-Sainio P, Rötter RP, Skjelvåg AO (2012) Shifts in comparative advantages for maize, oat, and wheat cropping under climate change in Europe. Food Addit Contam 29:1514–1526

Gleckler PJ, Taylor KE, Doutriaux C (2008) Performance metrics for climate models. J Geophys Res 113

Jensen HS, Andersen FØ (1992) Importance of temperature, nitrate and pH for phosphorus from aerobic sediments of four shallow, eutrophic lakes. Limnol Oceanogr 37:577–589

Jentsch A, Kreyling J, Beierkuhnlein C (2007) A new generation of climate-change experiments: events, not trends. Front Ecol Environ 5(7):365–374

Jørgensen SE (1995) State of the art of ecological modelling in limnology. Ecol Model 78:101–115

Karlsson IB, Sonnenborg TO, Jensen KH, Refsgaard JC (2013) Evaluating the influence of long term historical climate change on catchment hydrology—using drought and flood indices. Hydrol Earth Syst Sci Discuss 10:2373–2428

Klemes V (1986) Operational testing of hydrological simulation models. Hydrol Sci J 31:13–24

Liboriussen L, Lauridsen TL, Søndergaard M, Landkildehus F, Larsen SE, Jeppesen E (2011) Effects of warming and nutrients on sediment community respiration in shallow lakes: an outdoor mesocosm experiment. Freshw Biol 56(3):437–447

Madsen H (2000) Automatic calibration of a conceptual rainfall-runoff model using multiple objectives. J Hydrol 235:276–288

Meerhoff M, Teixeira-de Mello F, Kruk C, Alonso C, González-Bergonzoni I, Pacheco JP, Lacerot G, Arim M, Beklioğlu M, Bruce S, Goyenola G, Iglesias C, Mazzeo N, Kosten S, Jeppesen E (2012) Environmental warming in shallow lakes: a review of potential changes in community structure as evidenced from space-for-time substitution approaches. Adv Ecol Res 46:259–349

Murphy JM, Booth BBB, Collins M, Harris GR, Sexton DMH et al (2007) A methodology for probabilistic predictions of regional climate change from perturbed physics ensembles. Phil Trans R Soc A Math Phys Eng Sci 365:1993–2028

Olsen J, Anderson NJ, Knudsen MF (2012) Variability of the North Atlantic Oscillation over the past 5,200 years. Nat Geosci 5:808–812

Oreskes N, Shrader-Frechette K, Belitz K (1994) Verification, validation and confirmation of numerical models in the earth sciences. Science 264:641–646

Palosuo T, Kersebaum KC, Angulo C, Hlavinka P, Moriondo M, Olesen JE, Patil R, Ruget F, Rumbaur C, Takáč J, Trnka M, Bindi M, Caldag B, Ewert F, Ferrise R, Mirschel W, Saylan L, Šiška B, Rötter R (2011) Simulation of winter wheat yield and its variability in different climates of Europe. A comparison of eight crop growth models. Eur J Agron 35:103–114

Popper KR (1959) The logic of scientific discovery. Hutchingson & Co, London

Refsgaard JC, Henriksen HJ (2004) Modelling guidelines—terminology and guiding principles. Adv Water Resour 27(1):71–82

Refsgaard JC, Knudsen J (1996) Operational validation and intercomparison of different types of hydrological models. Water Resour Res 32(7):2189–2202

Refsgaard JC, Arnbjerg-Nielsen K, Drews M, Halsnæs K, Jeppesen E, Madsen H, Markandya A, Olesen JE, Porter JR, Christensen JH (2013) The role of uncertainty in climate change adaptation strategies—a Danish water management example. Mitig Adapt Strateg Glob Chang 18:337–359

Rötter R, Carter TR, Olesen JE, Porter JR (2011) Crop-climate models need an overhaul. Nat Clim Chang 1:175–177

Schlesinger S, Crosbie RE, Gagné RE, Innis GS, Lalwani CS, Loch J, Sylvester J, Wright RD, Kheir N, Bartos D (1979) Terminology for model credibility. SCS Technical Committee on Model Credibility. Simulation 32(3):103–104

Seibert J (2003) Reliability of model predictions outside calibration conditions. Nord Hydrol 34(5):477–492

Singh R, Wagener T, van Werkhoven K, Mann ME, Crane R (2011) A trading-space-for-time approach to probabilistic continuous streamflow predictions in a changing climate—accounting for changing watershed behaviour. Hydrol Earth Syst Sci 15(11):3591–3601

Sunyer MA, Madsen H, Rosbjerg D, Arnbjerg-Nielsen K (2013) Regional interdependency of precipitation indices across Denmark in two ensembles of high resolution RCMs. J Clim. doi:10.1175/JCLI-D-12-00707.1

Swayne D, Lam D, MacKay M, Rouse W, Schertzer W (2005) Assessment of the interaction between the Canadian Regional Climate Model and lake thermal-hydrodynamic models. Environ Model Softw 20:1505–1513

Taylor KE, Stouffer RJ, Meehl GA (2012) An overview of CMIP5 and the experiment design. Bull Am Meteorol Soc 93:485–498

Tebaldi C, Smith R, Nychka D, Mearns L (2005) Quantifying uncertainty in projections of regional climate change: a Bayesian approach to the analysis of multi-model ensembles. J Clim 18:1524–1540

Teutschbein C, Seibert J (2012) Is bias correction of Regional Climate Model (RCM) simulations possible for non-stationary conditions? Hydrol Earth Syst Sci Discuss 9:12765–12795

Trolle D, Hamilton DP, Pilditch CA, Duggan IC, Jeppesen E (2011) Predicting the effects of climate change on trophic status of three morphologically varying lakes: Implications for lake restoration and management. Environ Model Softw 26:354–370

Van Der Linden P, Mitchell JFB (eds) (2009) ENSEMBLES: climate change and its impacts: summary of research and results from the ENSEMBLES project. Met Office Hadley Centre, Exeter

Van Steenbergen N, Willems P (2012) Method for testing the accuracy of rainfall-runoff models in predicting peak flow changes due to rainfall changes, in a climate changing context. J Hydrol 414–415:425–434

Velázquez JA, Schmid J, Ricard S, Muerth MJ, Gauvin St-Denis B, Minville M, Chaumont D, Caya D, Ludwig R, Turcotte R (2012) An ensemble approach to assess hydrological models' contribution to uncertainties in the analysis of climate change impact on water resources. HESSD, 9, 7441–7474

Wagener T, Sivapalan M, Troch PA, McGlynn BL, Harman CJ, Gupta HV, Kumar P, Rao PSC, Basu NB, Wilson JS (2010) The future of hydrology: an evolving science for a changing world. Water Resour Res 46, W05301

Wilby RL, Harris I (2006) A framework for assessing uncertainties in climate change impacts: Low-flow scenarios for the River Thames, UK. Water Resour Res 42:W02419

Wolf J, Evans LG, Semenov MA, Eckersteen H, Iglesias A (1996) Comparison of wheat simulation models under climate change. I. Model calibration and sensitivity analyses. Clim Res 7:253–270